

# PATENT ABSTRACTS OF JAPAN

(11)Publication number : 2000-078530

(43)Date of publication of application : 14.03.2000

(51)Int.Cl.

H04N 5/928  
G06F 17/30  
G10L 15/00

(21)Application number : 10-257668

(71)Applicant : NEC CORP

(22)Date of filing : 28.08.1998

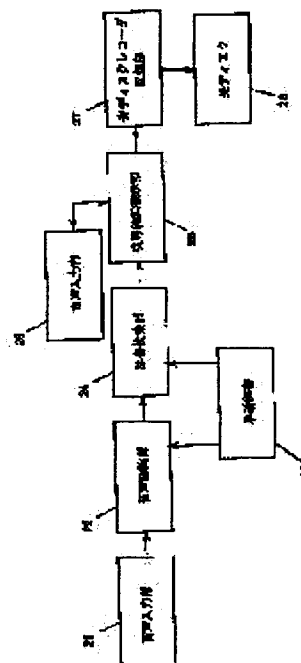
(72)Inventor : MATSUI TSUTOMU

## (54) INFORMATION RECORDER, INFORMATION RECORDING METHOD AND RECORDING MEDIUM

### (57)Abstract:

PROBLEM TO BE SOLVED: To input a keyword in voice and to record it in relation to video/audio data.

SOLUTION: A voice recognition section 22 recognizes a word with respect to a voice entered by a voice entry section 21 by using a word dictionary 23 and gives the word to a dictionary retrieval section 24. The dictionary retrieval section 24 retrieves information of this word from the word dictionary 23 and gives the result of retrieval to a retrieval result display section 25. The retrieval result display section 25 displays plurality of retrieval results sequentially, a voice input section 26 confirms one retrieval result and gives it to an optical disk recorder recording section 27. The optical disk recorder recording section 27 records the confirmed retrieval result to an optical disk 28 in relation to video/audio data.



## LEGAL STATUS

[Date of request for examination] 28.08.1998

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number] 3166725

[Date of registration] 09.03.2001

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号  
特開2000-78530  
(P2000-78530A)

(43) 公開日 平成12年3月14日 (2000.3.14)

(51) Int.Cl. <sup>7</sup>	識別記号	F I	テーマコード* (参考)
H 0 4 N 5/928		H 0 4 N 5/92	E 5 B 0 7 5
G 0 6 F 17/30		G 1 0 L 3/00	5 5 1 G 5 C 0 5 3
G 1 0 L 15/00		G 0 6 F 15/40	3 7 0 D 5 D 0 1 5
		15/401	3 1 0 C

審査請求 有 請求項の数 8 F D (全 6 頁)

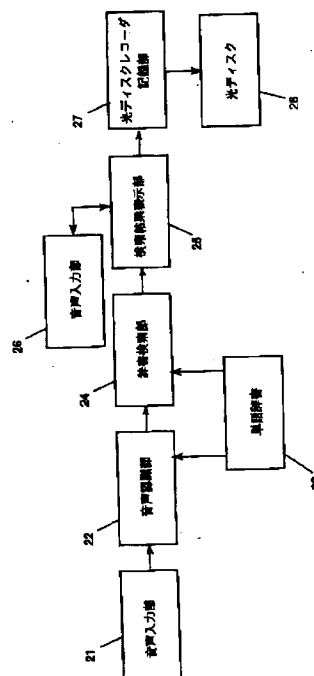
(21) 出願番号	特願平10-257668	(71) 出願人	000004237 日本電気株式会社 東京都港区芝五丁目7番1号
(22) 出願日	平成10年8月28日 (1998.8.28)	(72) 発明者	松井 勉 東京都港区芝五丁目7番1号 日本電気株式会社内
		(74) 代理人	100097113 弁理士 堀 城之 Fターム(参考) 5B075 ND08 ND14 NK13 NK25 PP02 PP03 PP13 PP22 PQ02 PQ46 PQ48 PQ66 UU29 UU35 5C053 FA06 FA14 FA25 HA29 JA01 JA23 JA30 LA01 LA20 5D015 AA04 BB01 KK02

(54) 【発明の名称】 情報記録装置および情報記録方法、並びに記録媒体

(57) 【要約】

【課題】 音声によりキーワードを入力し、映像／音声に関連付けて記録する。

【解決手段】 音声入力部21より入力された音声に対応する単語が、音声認識部22により単語辞書23を用いて認識され、辞書検索部24に供給される。辞書検索部24は、この単語の情報を単語辞書23から検索し、検索結果を検索結果表示部25に供給する。検索結果表示部25は、検索結果が複数ある場合、順番に表示し、音声入力部26により1つの検索結果が確定され、光ディスクレコーダ記録部27に供給する。光ディスクレコーダ記録部27は、確定された検索結果を映像／音声に関連づけて光ディスク28に記録する。



## 【特許請求の範囲】

【請求項 1】 映像を入力する映像入力手段と、  
 音声を入力する音声入力手段と、  
 前記音声入力手段によって入力された前記音声を認識し、前記音声に対応する単語に変換する音声認識手段と、  
 前記音声認識手段によって認識され、変換された前記単語に対応するコードを生成する生成手段と、  
 前記映像入力手段によって入力された前記映像と、前記生成手段によって生成された前記単語に対応する前記コードを対応付けて、所定の記録媒体に記録する記録手段と、  
 を備えることを特徴とする情報記録装置。

【請求項 2】 前記記録手段は、前記音声入力手段によって入力された前記音声を、前記映像及び前記コードとともに前記記録媒体に記録することを特徴とする請求項 1 に記載の情報記録装置。

【請求項 3】 前記単語と前記単語に関する情報を関連付けて記憶する記憶手段をさらに備え、  
 前記音声認識手段は、前記記憶手段に記憶されている前記単語に基づいて、前記音声を前記単語に変換することを特徴とする請求項 1 に記載の情報記録装置。

【請求項 4】 前記単語と前記単語に関する情報を関連付けて記憶する記憶手段と、  
 前記音声認識手段によって認識された前記単語の情報を、前記記憶手段から検索する検索手段と、  
 前記検索手段によって検索された前記単語の情報を表示する表示手段とをさらに備えることを特徴とする請求項 1 に記載の情報記録装置。

【請求項 5】 前記単語と前記単語に関する情報を関連付けて記憶する記憶手段と、  
 前記音声認識手段によって認識された前記単語を、前記記憶手段から検索する検索手段と、  
 前記検索手段によって検索された前記単語の情報を表示する表示手段と、  
 前記検索手段によって複数の情報が検索されたとき、前記表示手段に複数の情報を順次表示させ、複数の情報から 1 つの情報を選択する選択手段とをさらに備え、  
 前記記録手段は、前記選択手段によって選択された前記情報に対応する単語のコードを前記記録媒体に記録することを特徴とする請求項 1 に記載の情報記録装置。

【請求項 6】 前記コードと論理符号から検索式を生成する検索式生成手段をさらに備え、  
 前記記録手段は、前記コードおよび前記検索式と前記映像を関連付けて、前記記録媒体に記録することを特徴とする請求項 1 に記載の情報記録装置。

【請求項 7】 映像を入力する映像入力ステップと、  
 音声を入力する音声入力ステップと、  
 前記音声入力ステップにおいて入力された前記音声を認識し、前記音声に対応する単語に変換する音声認識ス

ップと、  
 前記音声認識ステップにおいて認識され、変換された前記単語に対応するコードを生成する生成ステップと、  
 前記映像入力ステップにおいて入力された前記映像と、  
 前記生成ステップにおいて生成された前記単語に対応する前記コードを対応付けて、所定の記録媒体に記録する記録ステップとを備えることを特徴とする情報記録方法。

【請求項 8】 請求項 7 に記載の情報記録方法を実行可能なプログラムを記録したことを特徴とする記録媒体。

## 【発明の詳細な説明】

## 【0001】

【発明の属する技術分野】本発明は、情報記録装置および情報記録方法、並びに記録媒体に関し、特に、音声入力からキーワードを生成し、映像とキーワードを関連付けて記録することにより、キーワードによる映像の検索を可能とする情報記録装置および情報記録方法、並びに記録媒体に関する。

## 【0002】

【従来の技術】従来の VTR 型ビデオカメラでは、カメラ部分に RET (リターン) キーが用意されており、撮影中に編集の「きっかけ」となる部分についてのマークを入力することができるようになされている。従って、このマークを入力することにより、後で編集を行うときに、マークを入力した部分を簡単に呼び出すことができ、編集作業を容易にすることができる。

## 【0003】

【発明が解決しようとする課題】しかしながら、例えば、手術中に、手術の状況をビデオカメラ等で撮影するような場合、医師がカメラ操作を行い、マークを入力することは困難である。仮に、助手に依頼するにしても、医師の思った通りにマークを入力することは困難である課題があった。

【0004】そこで、できれば音声入力によって、自動的に編集の「きっかけ」、又は手術中の大切な画面取りのタイミングを取ることが望まれる。

【0005】その場合、キーとなるワードが予め登録されたワードでないと、自動キーワード検索（音声入力の際に予め登録したワードと照合すること）を行うことができなくなる。手術中は手術に専念する必要があることから、音声入力により自動的に、キーワードをメタデータ（追加データ）として、映像／音声に加えてレコーダに入力する必要がある。

【0006】本発明はこのような状況に鑑みてなされたものであり、ビデオカメラの撮影中に、音声入力によりキーワードを入力し、後でキーワード検索が可能のように、映像／音声にキーワードを付加して記録することができるようにするものである。

## 【0007】

【課題を解決するための手段】請求項 1 に記載の情報記

録装置は、映像を入力する映像入力手段と、音声を入力する音声入力手段と、音声入力手段によって入力された音声を認識し、音声に対応する単語に変換する音声認識手段と、音声認識手段によって認識され、変換された単語に対応するコードを生成する生成手段と、映像入力手段によって入力された映像と、生成手段によって生成された単語に対応するコードを対応付けて、所定の記録媒体に記録する記録手段とを備えることを特徴とする。また、記録手段は、音声入力手段によって入力された音声を、映像及びコードとともに記録媒体に記録するようにすることができる。また、単語と単語に関する情報を関連付けて記憶する記憶手段をさらに設けるようにし、音声認識手段は、記憶手段に記憶されている単語に基づいて、音声を単語に変換するようにすることができる。また、単語と単語に関する情報を関連付けて記憶する記憶手段と、音声認識手段によって認識された単語の情報を、記憶手段から検索する検索手段と、検索手段によって検索された単語の情報を表示する表示手段とをさらに設けるようにすることができる。また、単語と単語に関する情報を関連付けて記憶する記憶手段と、音声認識手段によって認識された単語を、記憶手段から検索する検索手段と、検索手段によって検索された単語の情報を表示する表示手段と、検索手段によって複数の情報が検索されたとき、表示手段に複数の情報を順次表示させ、複数の情報から1つの情報を選択する選択手段とをさらに設けるようにし、記録手段は、選択手段によって選択された情報に対応する単語のコードを記録媒体に記録するようにすることができる。また、コードと論理符号から検索式を生成する検索式生成手段をさらに設けるようにし、記録手段は、コードおよび検索式と映像を関連付けて、記録媒体に記録するようにすることができる。請求項7に記載の情報記録方法は、映像を入力する映像入力ステップと、音声を入力する音声入力ステップと、音声入力ステップにおいて入力された音声を認識し、音声に対応する単語に変換する音声認識ステップと、音声認識ステップにおいて認識され、変換された単語に対応するコードを生成する生成ステップと、映像入力ステップにおいて入力された映像と、生成ステップにおいて生成された単語に対応するコードを対応付けて、所定の記録媒体に記録する記録ステップとを備えることを特徴とする。請求項8に記載の記録媒体は、請求項7に記載の情報記録方法を実行可能なプログラムを記録したことを特徴とする。本発明に係る情報記録装置および情報記録方法、並びに記録媒体においては、映像を入力し、音声を入力し、入力された音声を認識し、音声に対応する単語に変換し、音声認識され、変換された単語に対応するコードを生成し、映像とコードを対応付けて、所定の記録媒体に記録する。

【0008】

【発明の実施の形態】図1は、本発明の情報記録装置を

応用した光ディスクカメラ装置の一実施の形態の構成例を示す図である。図1は、医用のビデオ取りで用いられ、天井の照明の中心に取り付けられた様子を示している。光ディスクカメラ装置1（以下、適宜カメラと略記する）の操作は、医師の補助スタッフ等によって行われるものとする。

【0009】光ディスクカメラ装置1は、被写体からの光を集光し、被写体の映像を形成するレンズ部11と、周囲の音や医師が発音したキーワード等を集音するマイクロフォン13と、レンズ部11において形成された映像と、マイクロフォン13によって集音された音声を記録するカメラ部12より構成されている。

【0010】また、光ディスクカメラ装置1には、Z軸周りに光ディスクカメラ装置1を回転させるためのドライバ14と、光ディスクカメラ装置1をY軸周りに回転させるためのドライバ15を有しており、所定の方角に光ディスクカメラ装置1を向けることができるようになっている。

【0011】光ディスクカメラ装置1は、手術等の映像を撮影している時に、音声入力された音声を認識し、予め単語と単語の情報を登録した単語辞書23に照らし合わせて該当する単語を抽出し、それをキーワードとして、後でキーワード検索が可能のように、ビデオ記録／音声記録のフレーム若しくはフィールド内部のシステムデータ（図3）に、追加データ（メタデータ）として挿入するようになっている。これにより、撮影後に、キーワード検索により、目的とする映像／音声が記録されたフレームを引き出すことができる。

【0012】また、キーワードと論理符号（AND/OR）から検索式を生成し、それをシステムデータフィールドに記録するようにすることもできる。例えば、心臓／肺／バイパス／手術等の各キーワードを、AND/ORによって連結することにより検索式を生成し、それをシステムデータフィールドに記録しておくことで、この検索式を指定することにより、特定のビデオ映像を引き出すことができる。

【0013】図2は、図1に示した光ディスクカメラ装置1の内部の電氣的な構成例を示すブロック図である。

【0014】同図に示すように、光ディスクカメラ装置1は、音声を入力し、この音声入力を所定の音声データに変換する音声入力部21と、音声入力部21によって変換された音声データを音声認識する音声認識部22と、音声認識部22によって認識された音声認識結果の単語について、辞書検索を行なう辞書検索部24と、辞書検索部24によって検索された単語の情報をユーザに対して表示する検索結果表示部25とを有している。これらの音声入力部21、音声認識部22、辞書検索部24、検索結果表示部25は、辞書検索装置を構成している。

【0015】また、光ディスクカメラ装置1は、音声認

識部22が、音声入力部21より供給される音声データの音声認識に使用する各単語の音声認識用テンプレートと、辞書検索部24が検索する各単語の情報を、同一の単語エントリに登録してある単語辞書23を有している。ここで、音声入力部21及び音声入力部26は、図1におけるマイクロフォン13を構成している。

【0016】音声認識のためには、予め使用する言語を登録しておき、この登録した言語のどれに一番近いかを判別した方が高速に処理できる。このため、同一音声に対して、複数（例えば、2、3）の選択枝が発生した場合、どれかに確定する必要がある。この確定を音声入力部26から音声により入力するようにしている。

【0017】音声入力部26は、検索結果表示部25により音声認識の結果として表示された複数の認識結果の候補の選択枝に対して、YES又はNO、或いは、OK又はNGの判定を音声により入力するためのものである。このときは、音声入力部21から入力された音声は無視される。従って、音声入力部26は、YES/NO、OK/NGを認識する簡単な音声認識回路を備えているものとする。

【0018】例えば、「ishi」と発音し、音声入力部21より入力すると、音声認識部22は、単語辞書23から、「いし」という単語を認識する。次に、辞書検索部24は、「いし」という単語について、単語辞書23から、「医師」、「意志」、「石」、「縊死」等を検索する。そして、検索結果が順番に、検索結果表示部25により表示される。これに対して、音声入力部26より、YES若しくはNO、又はOK若しくはNGを音声によって入力することにより、複数の検索結果から、希望の単語を選択し、確定することができる。

【0019】図3は、ビデオ／オーディオデータが記録される光ディスク28のフレームのフォーマットの例を示している。各フレームは、同期をとるための同期部と、フレームの順番を示すIDが格納されるシンクブロックIDと、ビデオデータ、及びオーディオデータより構成される。また、システムデータが、ビデオデータの中に640バイト（B）（＝320＊2）、オーディオデータの中に40バイト（＝20＊2）、合わせて680バイト（B）／フィールド、即ち、1360バイト／フレーム入る。外符号訂正、内符号訂正は、それぞれ誤り訂正符号である。

【0020】ビデオデータのシステムデータには、音声入力されたキーワードが記録され、オーディオデータのシステムデータには、キーワード、検索式、後述する「YES/NO判定」に至った経過、ビデオデータのシステムデータに入り切らなかったデータ等が記録される。

【0021】従って、1フレームあたり、2バイト文字（全角）の場合、680字、1バイト文字（半角）の場合、1360文字を、システムデータ中にメタデータと

して自動記録することが可能となる。このとき、音声認識された音声の中のキーワードを抽出し、キーワードだけを記録するようにすると、結果として使用メモリ数を圧縮して記録することが可能である。

【0022】次に、その動作について説明する。医師が手術中に、助手が光ディスクカメラ装置1を操作し、手術の模様を撮影する。撮影中に、医師及び助手等が光ディスクカメラ装置1に記録したい音声を音声入力部21より入力する。

【0023】音声入力部21に入力された音声は、デジタルの音声データに変換された後、音声認識部22に供給される。音声認識部22は、単語辞書23を用いて、音声入力部21より供給された音声データに対して音声認識を行い、認識結果としての単語を、辞書検索部24に供給する。

【0024】辞書検索部24は、音声認識部22より供給された認識結果としての単語について、単語辞書23を用いて辞書検索を行う。そして、検索結果としての単語の情報、例えば、日本語の場合、漢字などの情報を、また、単語が複数検索された場合は、同音異義語などの情報を、さらに音声認識があいまいな場合、複数の検索結果を音声認識結果の候補として、検索結果表示部25に供給する。

【0025】検索結果表示部25は、辞書検索部24から供給された単語の情報を表示する。このとき、音声認識された単語が複数ある場合、単語の情報を1つずつ順番に表示し、その正否が指示されるのを待つ。

【0026】例えば、「YES/NO」、又は「OK/NG」という単語が、音声入力部26から入力されるのを待つ。音声入力部26は、簡単な音声認識機能を有しており、「YES」又は「OK」が入力されたとき、検索結果表示部25に対して、今、表示されている単語の情報が正しいことを示す信号を供給する。この信号を受けた検索結果表示部25は、いま、表示している単語が正しい認識結果であるとみなし、その単語をカメラ部12の光ディスクレコーダ記録部27に供給する。

【0027】一方、音声入力部26は、「NO」又は「NG」が入力されたとき、検索結果表示部25に対して、いま表示されている単語の情報が正しくないことを示す信号を供給する。この信号を受けた検索結果表示部25は、いま表示している単語が正しい認識結果ではないとみなし、次の単語を表示する。そして、再度、その正否を指示する「YES/NO」という単語が、音声により、音声入力部26から入力されるのを待つ。

【0028】音声入力部21より、例えば、いつ／どこで／誰が／どのような手術／を行ったかを入力する。具体的には、音声により、「せんきゅうひやくきゅうじゅうはちねんはちがつにじゅういちにち、じゅうじぜろぜろふんからじゅうにじぜろぜろふん、けいおうだいがくいがくぶしなのちょうでこばやしこうしがしつとう

し、ふじのきょうじゅがたちあう、ずがいこつせいけいしゅじゅつ、ぶらすちつくそうにゅう、かんじゃめい、XXXX、おとこ、にじゅうごさい」と入力すると、音声認識部 21 は、入力された音声に対応する単語を、単語辞書 23 から抽出し、辞書検索部 24 に供給する。辞書検索部 24 は、音声認識部 22 によって認識された単語について、単語辞書 23 からその詳細な情報を検索する。

【0029】例えば、「けいおうだいがく」に対しては、単語辞書 23 に予め登録されている「慶応大学」が検索される。検索結果が複数ある場合には、上述したように、複数の検索結果を順次表示し、その中から 1 つを音声により選択する。同様にして、他の単語に対しても、単語辞書 23 から検索し、対応する情報を抽出する。このようにして、音声入力された音声に対応するキーワードが確定される。

【0030】上記例の場合、音声認識の結果、次のようなキーワードが確定される。即ち、「082198/1000-1200/慶応大学医学部/信濃町/小林講師/藤野教授/頭蓋骨成形/プラスチック挿入/患者名：\*\*\*\*/男/25 歳」。

【0031】ここで、各キーワードをアルファベット (A、B、C、D、E、F、G、H、I、J、K) 又は番号 (1、2、3、4、5、6、7、8、9、10、11) で置き換え、例えば、 $((A+B)*E*I)$  等の検索式を作成し、上記キーワードとともに、光ディスクレコーダ記録部 27 に供給するようにしてもよい。上記検索式において、「+」は論理和を表し、「\*」は論理積を表している。

【0032】光ディスクレコーダ記録部 27 は、検索結果表示部 25 より供給されたキーワード、即ち、1 又は複数のキーワードに対応するコードを、図 3 に示したシステムデータの中に書き込む。また、検索結果表示部 25 より、検索式が供給された場合、上記キーワードの後に、検索式も書き込む。

【0033】このとき、検索結果表示部 25 は、上述したように、各映像/フレーム (静止画) ごとに、後でキーワード検索が可能のように、音声入力された 1 又は複数のキーワード (単語) と、必要に応じて検索式を、光ディスクレコーダ記録部 27 に供給する。例えば、上記例のように、「082198/1000-1200/慶応大学医学部/信濃町/小林講師/藤野教授/頭蓋骨成形/プラスチック挿入/患者名：\*\*\*\*/男/25 歳」というキーワードと、検索式  $((A+B)*E*I)$  を光ディスクレコーダ記録部 27 に供給する。

【0034】光ディスクレコーダ記録部 27 は、図 3 に示したフレームのビデオデータのシステムデータ、及びオーディオデータのシステムデータに、上記キーワードと検索式を記録する。

【0035】このようにして、映像を構成する各フレー

ム (静止画) 毎に、音声入力したキーワードを記録するとともに、必要に応じて、検索式を生成しながら記録することができる。

【0036】撮影後は、例えば、「082198/1000-1200/小林講師/患者名：\*\*\*\*/」を入力すると、カメラ部 12 の光ディスク 28 から、入力されたキーワードと同一のキーワードがシステムデータに記録されているフレームが検索され、該当する映像又は静止画が抽出される。該当する映像又は静止画が複数ある場合は、画面を分割した小画面等に多面表示される。ここで、所定の映像又は静止画を指示することにより、映像の場合、指示された映像が再生され、静止画の場合、画面一杯に表示される。

【0037】以上説明したように、撮影時に、音声入力された単語を認識し、その単語について自動的に辞書検索を行い、検索結果を、いま記録している映像や音声に関連づけて記録するので、後でキーワードに基づいた検索を行うことができ、編集作業が容易となり、高速に編集することができる。しかも、キーワード検索を介して容易にアーカイブすることができる。

【0038】また、音声によりキーワードを入力するので、手術中の医師などのように、手が放せない状況下でも使用することができる。

【0039】なお、上記実施の形態においては、本発明を医用カメラに応用する場合について説明したが、自動編集、自動アーカイブ、高速アーカイブ等にも本発明を適用することができる。

【0040】また、上記実施の形態においては、本発明を光ディスクカメラ装置に応用する場合について説明したが、デジタルビデオカメラ等、その他の撮像装置にも本発明を適用することができる。

【0041】また、上記実施の形態におけるフレームフォーマットは例であって、これに限定されるものではない。

【0042】

【発明の効果】以上の如く、本発明に係る情報記録装置および情報記録方法、並びに記録媒体によれば、映像と音声を入力し、入力された音声を認識して、音声を単語に変換し、音声認識され、変換された単語に対応するコードを生成し、映像とコードを対応付けて所定の記録媒体に記録するようにしたので、ビデオカメラ等で映像と音声を記録している最中に、音声入力によりキーワードを入力し、映像と音声にキーワードを付加して記録することができ、後でキーワードを用いて所望の映像と音声を検索することが可能となる。

【図面の簡単な説明】

【図 1】本発明の情報記録装置を応用した光ディスクカメラ装置の一実施の形態の構成例を示すブロック図である。

【図 2】図 1 の実施の形態の内部の電気的な構成例を示

\* 21, 26 音声入力部

22 音声認識部

23 單語辭書

- 24 辞書検索部

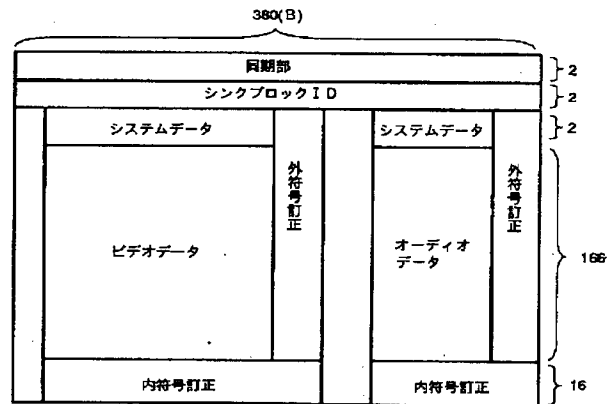
- ## 25 検索結果表示部

- 27 光ディスクレコ

- ## 28 光ディスク

\*

【図 3】



- 1 光ディスクカメラ装置  
11 レンズ部  
12 カメラ部  
13 マイクロフォン  
14, 15 ドライブ

【図 2】

